

Інформатика та системні науки (ІСН-2017)

УДК 004.658.3

ОПТИМІЗАЦІЯ МЕТОДУ ДОСТУПУ ДО ДАНИХ У РОЗПОДІЛЕНИХ СХОВИЩАХ

Б. В. Іваніщев, аспірант

Національний технічний університет України «Київський
політехнічний інститут імені Ігоря Сікорського»
callidus.iv@gmail.com

В статті розглядається оптимізація методу доступу до даних у розподілених сховищах за часом доступу. Пропонується новий алгоритм розміщення даних.

Ivanishchev B. V. Optimization of data access method for distributed storages. In the article are discussed optimization of data access method in distributed storages. A new algorithm of the solution of problem of data placement is proposed.

Ключовые слова: ОПТИМІЗАЦІЯ, ЗАДАЧА РОЗМІЩЕННЯ, РОЗПОДІЛЕНІ СХОВИЩА ДАНИХ.

Keywords: OPTIMIZATION, PLACEMENT PROBLEM, DISTRIBUTED DATA STORAGEES.

У цій роботі пропонується оптимізація методу доступу до даних у розподілених сховищах за часом доступу. Під часом доступу розуміється час, який пройшов від моменту, коли користувач подав запит на отримання даних, до моменту, коли він ці данні отримав.

Запропонований спосіб заснований на новому алгоритмі розміщення даних по вузлах зберігання під час проведення процедури балансування даних. Цей алгоритм дозволяє розмістити дані фізично якомога ближче до користувачів, які найчастіше їх використовують. Це дозволяє зменшити час доступу до даних зацікавлених у них користувачів.

Розподілене сховище даних може бути описане графом G , який включає множину пристроїв зберігання та множину зв'язків між ними

$$\begin{aligned} G &= (V, E, P), \\ E &\subseteq V \times V, \\ P &: E \rightarrow N, \end{aligned} \quad (1)$$

$$\forall v, \omega \in V \Rightarrow P(v, \omega) = 0, \text{ якщо } (v, \omega) \notin E,$$

$$\forall v, \omega \in V \Rightarrow P(v, \omega) = P(\omega, v),$$

де V – множина пристроїв зберігання даних, E – множина зв'язків між пристроями зберігання даних, P – вагова функція графу (час необхідний на переміщення об'єкту одиничного розміру за допомогою зв'язку або швидкість передачі даних за допомогою цього зв'язку).

У статті [1] описаний існуючий алгоритм балансування навантаження для розподіленої обчислювальної системи. Згідно нього задача пошуку збалансованого розміщення даних полягає в пошуку збалансованого потоку, який знаходиться розв'язанням наступної задачі лінійного програмування

$$\begin{aligned} c_i + \sum_{(j,i) \in E} x_{ji} - \sum_{(i,j) \in E} x_{ij} &\leq w, i \in V, \\ x_{ij} &\geq 0, (i, j) \in E, \\ \sum_{(j,i) \in E} x_{ji} &\rightarrow \min, \end{aligned} \quad (2)$$

де c_i – завантаженість i вузла зберігання даних, w – середня завантаженість пристроїв зберігання, x_{ij} – деякий об'єм даних, які треба перемістити з i пристрою зберігання на j пристрій зберігання.

Оптимізований алгоритм враховує час передачі даних від пристрою зберігання, до вузлів, з яких відбувається доступ зацікавлених користувачів. Для цього для кожного елементу даних з множини даних в сховищі O задано множину пристроїв зберігання, з яких відбувався доступ до цього елементу, $I_o, o \in O, I_o \subseteq V$.

Далі для кожної пари елемент даних – пристрій зберігання $(o, v), o \in O, v \in V$ можна визначити показник ефективності розміщення елементу даних o на пристрої зберігання v за формулою

$$a_{ov} = d_o \cdot \frac{1}{|I_o|} \cdot \sum_{j \in I_o} p_{vj}, p_{vj} \in P, (v, j) \in E, \quad (3)$$

де d_o – розмір елементу даних, p_{vj} – час необхідний на переміщення об'єкту одиничного розміру між пристроями зберігання v та j .

Цей показник визначає середній час передачі об'єкту від пристрою, на якому він зберігається, до пристроїв, з яких відбувався та потенційно може відбутися доступ до цього об'єкту.

За основу алгоритму, що розробляється, візьмемо існуючий алгоритм розміщення. Нехай $x_{ov}=1$, якщо елемент даних o розміщений на пристрої зберігання v , та $x_{ov}=0$, якщо не розміщений. Тоді задача пошуку збалансованого розміщення полягає в пошуку усіх величин $x_{ov}, o \in O, v \in V$, які знаходяться як розв'язок наступної задачі лінійного програмування

$$\begin{aligned} \sum_{o \in O, v \in V} d_o \cdot x_{ov} &\leq w, \\ f(o, v) = \sum_{o \in O, v \in V} a_{ov} \cdot x_{ov} &\rightarrow \min, \end{aligned} \quad (4)$$

У цій роботі був запропонований новий алгоритм розв'язання задачі розміщення даних, який урахує час передачі даних від вузла, на якому вони зберігаються, до вузлів, з яких відбувається доступ до них. Використання цього алгоритму дозволяє оптимізувати доступ до даних у розподілених сховищах за часом доступу.

Литература

1. Афраймович Л. Г. Навчально-методичний посібник з курсу «Моделі і методи ефективного використання розподілених обчислювальних систем» при вивченні теми «Завдання статичного балансування»: Навчально-методичний посібник. — Ніжній Новгород, 2012. — 13 с.
2. Привалов А. Н. Розробка балансувальника навантаження в гібридних мережах / А. Н. Привалов, А. К. Клепіков // Известія Тульського державного університету. – 2014. – №9–2. – С. 99–107.